



Skill forecasting from different wind power ensemble prediction methods

Pierre Pinson, Henrik A. Nielsen, Henrik Madsen, Georges Kariniotakis

► To cite this version:

Pierre Pinson, Henrik A. Nielsen, Henrik Madsen, Georges Kariniotakis. Skill forecasting from different wind power ensemble prediction methods. The Science of Making Torque from Wind, Aug 2007, Kongens Lyngby, Denmark. pp.012046, 10.1088/1742-6596/75/1/012046 . hal-00837982

HAL Id: hal-00837982

<https://hal-mines-paristech.archives-ouvertes.fr/hal-00837982>

Submitted on 24 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Skill forecasting from different wind power ensemble prediction methods

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2007 J. Phys.: Conf. Ser. 75 012046

(<http://iopscience.iop.org/1742-6596/75/1/012046>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 194.214.158.222

The article was downloaded on 24/06/2013 at 14:02

Please note that [terms and conditions apply](#).

Skill forecasting from different wind power ensemble prediction methods

Pierre Pinson¹, Henrik Aa. Nielsen¹, Henrik Madsen¹ and George Kariniotakis²

¹ Technical University of Denmark, Informatics and Mathematical Modelling, 2900 Kgs. Lyngby, Denmark

² Ecole des Mines de Paris, Center for Energy and Processes, 06904 Sophia Antipolis, France

E-mail: pp@imm.dtu.dk

Abstract. This paper presents an investigation on alternative approaches to the providing of uncertainty estimates associated to point predictions of wind generation. Focus is given to skill forecasts in the form of prediction risk indices, aiming at giving a comprehensive signal on the expected level of forecast uncertainty. Ensemble predictions of wind generation are used as input. A proposal for the definition of prediction risk indices is given. Such skill forecasts are based on the dispersion of ensemble members for a single prediction horizon, or over a set of successive look-ahead times. It is shown on the test case of a Danish offshore wind farm how prediction risk indices may be related to several levels of forecast uncertainty (and energy imbalances). Wind power ensemble predictions are derived from the transformation of ECMWF and NCEP ensembles of meteorological variables to power, as well as by a lagged average approach alternative. The ability of risk indices calculated from the various types of ensembles forecasts to resolve among situations with different levels of uncertainty is discussed.

1. Introduction

The large scale integration of wind generation capacities induces difficulties in the management of a power system. An additional challenge is to conciliate this deployment with the on-going deregulation of the European electricity markets. Increasing the value of wind generation through the improvement of prediction systems' performance is one of the priorities in wind energy research needs for the coming years [1]. Relevant prediction horizons are up to 72-hour ahead, justifying the choice for this forecast length in the present paper. A state of the art on wind power forecasting has been published by Giebel *et al* [2]. A large part of recent research works in wind power forecasting has concentrated on associating uncertainty estimates to point forecasts. Pinson and Kariniotakis [3] have described two complementary approaches that consist in providing forecast users with skill forecasts (commonly in the form of risk indices) or alternatively with probabilistic forecasts. For a thorough discussion on the probabilistic forecasting alternative, we refer to Pinson *et al* [4] and references therein. Here focus is given to the former type of uncertainty indicators.

It appears that low quality forecasts of wind generation are partly due to power prediction models, and partly to Numerical Weather Prediction (NWP) systems. Indeed, during some periods weather dynamics can be relatively more predictable, while at some other point in time they may prove to be unpredictable, and this regardless of the forecasting method employed.

Since power predictions are derived from nonlinear transformations of wind speed forecasts, the level of uncertainty in meteorological predictions may be amplified or dampened through this transformation. Providing forecast users with an *a priori* warning on the expected level of prediction uncertainty may allow them to develop alternative (and more or less risk averse) strategies. In an operational context, a skill forecast associated to a point prediction may be more easily understood than probabilistic forecasts. Also, skill forecasts are not directly related to a given prediction method: since they relate to an assessment of the inherent predictability of the weather dynamics, they are expected to provide an *a priori* warning whatever the considered single-valued forecast. Certain works in that direction are based on the definition of weather dynamics indicators [5]. More precisely, methods from synoptic climatology take advantage of measurements of wind speed and direction, as well as pressure, in order to classify local weather conditions. Consequently, [5] relates such typical weather patterns to different levels of forecast uncertainty. In contrast, ensemble forecasts of wind generation are used here as input for deriving skill forecasts. Such type of forecasts consists of a set of alternative prediction scenarios over a period of interest. Ensemble forecasts of meteorological variables are commonly produced by integrating the uncertainty in the computation of NWP [6]. Different types of meteorological ensemble predictions are considered, provided either by the European Centre for Medium-range Weather Forecasts (ECMWF, 51 members) [7, 8] or by the National Centre for Environmental Prediction (NCEP, 11 members) [9]. They are converted to ensemble predictions of wind power following the method described by Nielsen *et al* [10]. They will be referred as ECMWF-EPW and NCEP-EPW, respectively. A specificity of the conversion model in comparison with more classical power curve models is that it forces power forecasts to span the whole range of potential power values. Finally, lagged average ensembles, which consist in a set of forecasts with common lead time, but issued at different time origins, are used as a benchmark. They are here obtained by lagging wind power predictions produced with ECMWF control forecasts, and include 5 members. A full description of the various ensemble types is available in [11, Ch. 5].

The paper is structured as following. In a first stage, the methodology for skill forecasting is described, with emphasis on the definition of prediction risk indices and the way they should be related to forecast uncertainty. Then, the ability of the various risk indices to inform on expected uncertainty is evaluated and discussed by using the test case of a Danish offshore wind farm, over a period of 10 months. Concluding remarks end the paper.

2. Skill forecasts based on wind power ensembles

The methodology for skill forecasting dedicated to the wind power forecasting application is developed here. First, a definition of prediction risk indices is proposed. It is based on the dispersion of wind power ensembles over a single prediction horizon, or over a set of successive look-ahead times. It is then explained how such prediction risk indices may be used as skill forecasts, i.e. forecasts of the distributions of expected prediction errors. The relation between prediction risk indices and the level of prediction error is described with conditional probability diagrams.

2.1. Definition of prediction risk indices

Owing to the relation between the spread of ensemble members and the standard deviation of the errors described in e.g. [11], it is proposed here to define prediction risk indices as a measure of this ensemble spread. This measure is a continuous one, in contrast to some categorical measures¹ introduced in the meteorological literature, such as mode population [12] or ensemble statistical entropy [13]. Such choice is motivated by the conclusions from Grimit [14], stating

¹ The basic idea of categorical measures of ensemble spread consists in dividing the range of possible forecast values in several bins, and to count the numbers of ensemble members falling in each bin.

that continuous measures of ensemble spread are more appropriate if forecast's users have a continuous utility function². We assume that this is case or users of wind power predictions, either for the management or trading of wind generation.

Ensemble predictions of wind power issued at time t consist in a set of J alternative predictions $\hat{p}_{t+k/t}^{(j)}$ ($j = 1, \dots, J$) for any lead time $t + k$. The weighted standard deviation $\tilde{\sigma}_{t,k}$ of ensemble members is used as a measure of spread for that look-ahead time. $\tilde{\sigma}_{t,k}$ is given by

$$\tilde{\sigma}_{t,k} = \left[\frac{J}{J-1} \sum_{j=1}^J w_j \left(\hat{p}_{t+k/t}^{(j)} - \bar{p}_{t+k/t}^J \right)^2 \right]^{\frac{1}{2}} \quad (1)$$

such that the sum of the weights w_j totals 1, and with $\bar{p}_{t+k/t}^J$ the mean of the J alternative predictions for that lead time, that is

$$\bar{p}_{t+k/t}^J = \frac{1}{J} \sum_{j=1}^J \hat{p}_{t+k/t}^{(j)} \quad (2)$$

In equation (1), the weights may reflect the ability of the ensemble members to give an assessment of predictability. If considering for instance an algorithm that derives a best-guess forecast as a weighted average of the ensemble members, these weights could be directly used in the calculation of $\tilde{\sigma}_{t,k}$. A similar remark is valid if considering lagged average ensembles, for which the weights in the optimal combination of the alternative predictions are a function of their age.

Ensemble forecasts and power measures have different temporal resolution. Temporal interpolation is thus used in order for both of them to have a 15-minute resolution. Even though, the actual temporal resolution of ECMWF and NCEP meteorological ensembles is of six hours. In addition, weather predictability does not have an instantaneous nature: it is very unlikely that wind generation would be easily predictable for a given look-ahead time, and then highly unpredictable for the following one. Hence, it is envisaged here to estimate predictability over a time period. The use of the weighted standard deviation is generalized by computing the average of $\tilde{\sigma}_{t,k}$ over a set of consecutive horizons, from look-ahead time k_1 to k_2 . This average weighted standard deviation defines a *Normalized Prediction Risk Index*, abbreviated NPRI, calculated as

$$\text{NPRI}(k_1, k_2) := \frac{1}{k_2 - k_1 + 1} \sum_{i=k_1}^{k_2} \tilde{\sigma}_{t,i} \quad (3)$$

with $\tilde{\sigma}_{t,i}$ given by equation (1). In the following, NPRI_h denotes the prediction risk index calculated on a per-horizon basis (i.e. such that $k_1 = k_2$) while NPRI_d stands for periods of 24 hours (i.e. day 1, day 2, etc).

2.2. Relating NPRI and prediction errors

2.2.1. Considering prediction errors as energy imbalances Prediction errors are expressed in the form of energy imbalances since it is aimed at showing that NPRI can be used for informing on the level of expected prediction error over a certain period of time. Energy imbalances are defined here as the difference, in absolute value, between predicted and measured amounts of energy over a period of interest. Both measured and predicted amounts of energy are normalized

² The utility function for a forecast's user is introduced and further discussed by Pinson *et al* [15].

quantities, since power forecasts and measures are normalized values (by the nominal capacity P_n). The energy imbalance $d_{t+k_1}^{t+k_2}$ between lead time $t + k_1$ and lead time $t + k_2$ is

$$d_{t+k_1}^{t+k_2} = |E_{t+k_1}^{t+k_2} - \hat{E}_{t+k_1}^{t+k_2}| = t_r \sum_{i=k_1}^{k_2} |p_{t+i} - \hat{p}_{t+i/t}| \quad (4)$$

where $\hat{E}_{t+k_1}^{t+k_2}$ and $E_{t+k_1}^{t+k_2}$ are the predicted and measured quantities of energy over this period of time, while t_r is the temporal resolution of wind power predictions.

By calculating energy imbalances in absolute value, production surplus and shortage are similarly accounted for. Prediction risk indices are meant for estimating the expected level of uncertainty, but cannot give the sign of forecast errors. If considering a single look-ahead time, the normalized imbalance equals the prediction error in absolute value. And, for successive horizons, it is equivalent to the average absolute error over this time interval.

Prediction risk indices should provide information on the expected level of forecast uncertainty whatever the point prediction method considered. Therefore, we do not concentrate hereafter on the use of the best available point forecast of wind generation that can be derived from ensembles, i.e. given by the ensemble mean (or the weighted mean for lagged average ensembles) [11]. Instead, point predictions are produced as it is commonly done today, that is by applying a statistical power curve model to the control forecasts provided by meteorological offices, here from ECMWF or NCEP.

2.2.2. Conditional probability diagrams for relating NPRI to the level of expected prediction error Following discussion in [14], the relationship between prediction risk indices and level of prediction error is drawn from a probabilistic perspective. This proposal goes against the traditional approach consisting in fitting a linear regressor between measures quantifying the ensemble spread and predictor's skill, associated with a correlation coefficient assessing the strength of this relation (see [16, 17, 18] among others). The inconsistency of using the correlation coefficient has been discussed by Grimit and Mass [17]: considering it for measuring the strength of the relationship between the ensemble spread and the predictor's skill implicitly assumes a linear relation between these two variables, which is not true in practice³.

A possibility for expressing the relation between NPRI and related prediction error in a probabilistic manner is to use contingency tables, which give the probabilities of events defined by the occurrence of NPRI-range/error-range pairs. Such idea has been proposed first by Houtekamer [19] and consequently applied by Whitaker and Loughé [18]. Though, our choice goes for conditional probability diagrams similar to those used by Moore and Kleeman [20], which easily give a visual information on the relation between NPRI and prediction uncertainty levels.

Conditional probability diagrams summarize distributions of energy imbalances given NPRI values. The range of NPRI values is divided into categories defined as equally populated classes. This follows from the idea that it is not the value of NPRI by itself that tells if the situation is more or less uncertain, but more where this value is located in the climatological distribution of NPRI values [13, 18]. Also, considering equally populated classes of NPRI values allows us to compare skill forecasts made from ECMWF-EPW, NCEP-EPW or lagged average ensembles as input, independently of the range of their ensemble spread values. A similar reasoning applies for energy imbalances, which are normalized by their climatological value depending on the look-ahead period. This climatological value corresponds to the average imbalance over the

³ Actually, it has been shown for an ideal ensemble of infinite size that the spread-error correlation can be written analytically, as a function of the temporal variability of the ensemble spread [19]. In this model, the prediction error is in absolute value. For an infinite spread variability, this spread-error correlation asymptotes to 0.8 [18].

10-month evaluation period for each look-ahead period. When mentioning imbalance levels, they will indeed be relative and expressed in percentage of their climatological value. Thus, we will study how NPRI has the ability to tell if these imbalances are lower or higher than usual, independently of the global performance of the considered point prediction method.

3. Results

In this Section is evaluated the ability of prediction risk indices to differentiate between situations with low and high uncertainty depending on the use of ECMWF-EPW, NCEP-EPW, or lagged average ensembles as input. This study is for the test case of the Tunø Knob wind farm, located few kilometers off the east coast of Jutland in Denmark, with a nominal capacity P_n of 5MW. The period for which both meteorological and power data are available covers the first 10 months of 2003. The temporal resolution of 15 minutes for power measurements is chosen as the temporal resolution for the study.

3.1. Pointwise estimation of expected uncertainty

In a first stage, the ability of NPRI to inform on the level of prediction uncertainty when calculated for each look-ahead time is evaluated, for the three sets of wind power ensemble predictions. As explained in paragraph 2.1, NPRI_h corresponds to the weighted standard deviation of the ensemble members for a given look-ahead time. The weights for its calculation are set to $1/J$ for ECMWF-EPW and NCEP-EPW (where J the number of ensemble members equals 11 and 51, respectively). Alternatively, the weights given in table 1 are used for the case of the lagged averaged ensembles, following [11]. Owing to the limited amount of data available (only 300 series of wind power predictions over a 10-month evaluation period), and also for comparison with results of the second part of the study, NPRI and energy imbalance values are gathered for each day ahead.

Table 1. Weights used for calculating the NPRI when considering the lagged average ensembles. These weights are those that would also be used for combination of ensemble members in order to obtain an optimal single-valued prediction.

Forecast age [hours]	24	48	72	96
Weight	0.36	0.26	0.21	0.17

3.1.1. The NPRI_h ability to inform on the expected imbalance level Figure 1 gives the example of a conditional probability diagram for ECMWF-EPW for day 3 (i.e. for look-ahead times between 48- and 72-hour ahead). It takes the form of a set of boxplots, summarizing distributions of energy imbalances given the class of NPRI_h values. Boxplots are centred on the average NPRI_h values for each class. Five different NPRI classes are considered, for which the related empirical distributions of imbalances are composed by 1440 items each.

Evolution of mean values gives the general trend between NPRI_h and level of prediction error. There is a steady (and quasi linear) increase in the mean imbalance level when going from lowest to highest NPRI_h class. When NPRI_h values belong to the NPRI_h class 1, the average imbalance level equals 30% of the climatological one. Though, for class 5, this average imbalance is more than 5 times larger, reaching 155% of the climatological value. Using NPRI_h with ECMWF-EPW proves to be a possibility for resolving between situations with various levels of expected imbalances.

The most interesting information comes from the quoted quantiles of conditional probability distributions given the NPRI_h class, since it informs on a lower and an upper bound for expected

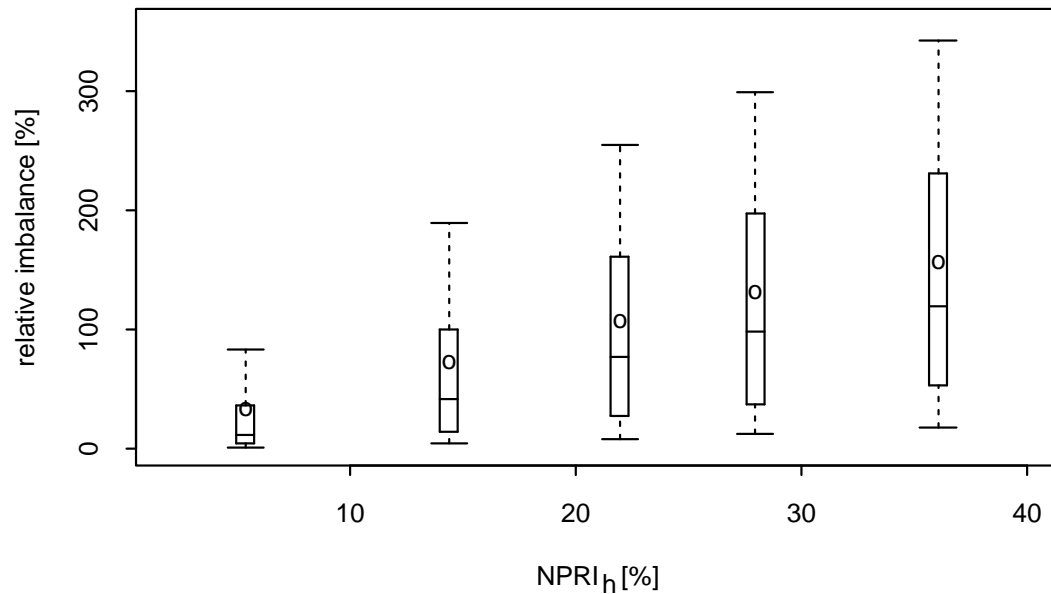


Figure 1. Conditional probability diagram giving the relation between NPRI_h and level of energy imbalance. NPRI_h values are calculated from ECMWF-EPW. Results are for day 3 (prediction horizons from 48 to 72-hour ahead). Empirical distributions are made up with 1440 elements. Boxplots give the 10% and 90% quantiles (lower and upper tips), the lower and upper quartiles (box bounds), the median (central line) and finally the mean (o).

imbalances. In figure 1, one sees for instance that if NPRI_h lies in the first class, then 90% of imbalances are below 90% of the climatological imbalance level (for the considered look-ahead time), while there is still a 10% probability that the level of imbalance exceeds 340% of the climatological imbalance value if the NPRI_h value belongs to class 5. The imbalance distributions become much wider when NPRI_h values are larger: the 10% quantiles are still close to zero, but the 90% ones get much higher. This upper bound on the expected imbalance level informs on the risk of relying on the provided wind power point prediction. From a risk aversion point of view, it would be preferable to make conservative decisions if NPRI_h values belong to class 5. Note that here, imbalances are relative to their climatological level in order to better illustrate the fact that prediction uncertainty is lower or higher than usual. Though, in an operational context, expected levels of imbalance could be expressed with physical units (e.g. in MWh or in percentage of the maximum possible generation over the time range).

The example of day 3 has been chosen for describing the relationship between NPRI_h calculated from ECMWF-EPW and level of imbalance in a probabilistic manner. The same kind of relation can be witnessed for the other days, or for the other types of ensembles. Though, that relationship may exhibit slightly different characteristics, which are studied in the following paragraph.

3.1.2. Comparing the cases for which NPRI_h is calculated with ECMWF-EPW, NCEP-EPW and the lagged average ensembles The relation between NPRI_h (for ECMWF-EPW) and imbalance

on a per-look-ahead time basis has been described above. Here, a comparison is made between the information content of the three types of ensemble predictions of wind generation considered. This comparison is possible for the first 3 days ahead. The specific case of day 1 is irrelevant since ECMWF-EPW are only available at the end of this first day in an operational context. Focus is given to day 2 (look-ahead times between 24 and 48-hour ahead) for highlighting the differences between the various wind power ensembles. Similar analyses were carried out for day 3, and the following comments are representative for the whole study.

Even if the NPRI_h values are not scattered in the same manner when considering NCEP-EPW, ECMWF-EPW, or lagged average ensembles, using classes of NPRI_h values enables to study the inherent ability of the various ensemble approaches to resolve between situations with lower and higher uncertainty. These categories of NPRI_h values permit to leave aside the problem of their distributions and to assess how their variations may inform on expected imbalance level. Therefore, when comparing the various approaches, we do not mention ranges of NPRI_h values, but only the NPRI_h class, numbered from 1 to 5. Table 2 gathers some of the quantiles ($r^{(\alpha)}$, with α the proportion) and the mean μ of the conditional probability distributions of imbalances, given the NPRI_h class, for the three types of ensemble predictions.

Table 2. Characteristics of conditional imbalance distributions given the NPRI_h class. $r^{(\alpha)}$ denotes the quantile with proportion α , while μ relates to the mean. Both NPRI_h and imbalance values are gathered over day 2 (look-ahead times between 24 and 48-hour ahead). Results are for ECMWF-EPW, NCEP-EPW, and the lagged average ensembles respectively. Empirical distributions are made up with 1440 elements. Imbalance values correspond to relative imbalances (in % of their climatological level).

(a) - ECMWF-EPW						
NPRI_h class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	1.3	2.9	10.2	28.7	68.9	27.7
2	4.2	15.2	42.7	89.6	176.7	70.5
3	10.8	30.1	78.4	157.7	277.7	113.7
4	15.2	44.3	113.5	202.7	298.3	137.4
5	16.0	47.5	116.4	223.5	333.6	150.7

(b) - NCEP-EPW						
NPRI_h class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	1.4	4.7	10.8	29.3	84.8	30.4
2	10.5	22.4	40.3	91.3	192.6	75.9
3	17.2	39.8	76.2	146.1	243.4	108.7
4	24.9	56.5	103.3	176.5	275.7	130.93
5	26.8	69.4	135.2	218.7	305.1	154.0

(c) - lagged average ensembles						
NPRI_h class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	1.4	3.7	13.0	40.2	99.8	40.0
2	5.1	20.0	57.3	117.0	223.5	89.1
3	9.2	28.9	88.4	171.6	264.4	114.9
4	11.2	35.3	101.8	192.6	290.2	127.6
5	8.7	27.1	76.7	187.8	323.5	128.3

The variability of the mean imbalance can be seen as a criterion for evaluating the ability

of the different approaches for dissociating between several levels of forecast uncertainty. That variability can be quantified by the ratio between the mean imbalances for NPRI_h class 5 and 1. This ratio is equal to 5.4, 5.1 and 3.2 for ECMWF-EPW, NCEP-EPW and the lagged average ensembles, respectively. ECMWF-EPW and NCEP-EPW have a higher differentiation ability (with a slight advantage for ECMWF-EPW), far better than that of the lagged average ensembles.

Then, focus is given to the quantiles of conditional imbalance distributions. The increase in the spread⁴ of these distributions when going from the first to the fifth class is more significant for ECMWF-EPW, followed by NCEP-EPW and the lagged average ensembles. This can also be seen as another criterion for stating that ECMWF-based ensembles better resolve among situations, since the variations in the range of expected imbalances are more pronounced. If looking separately at lower ($r^{(0.1)}$ and $r^{(0.25)}$) and upper ($r^{(0.75)}$ and $r^{(0.9)}$) quantiles, one sees that lower quantiles are more variable for NCEP-EPW while upper quantiles are more variable for ECMWF-EPW. The first one better resolves the low part of conditional imbalance distributions while the second better differentiates the upper part of these distributions. Therefore, if having a risk aversion point of view, NPRI_h used with ECMWF-EPW gives a more valuable information on the risk one may face when relying on the provided point prediction.

The increase in the mean imbalance depending on NPRI_h class is not as steady for the lagged average ensembles than for the others. In addition, the four lowest quantiles decrease between class 4 and 5. If the mean imbalance is higher for NPRI_h class 5, it is only because this NPRI_h class contains very large prediction errors. But, it also contains more low prediction errors than the fourth class. Note that this lagged averaging approach, even if less informative for skill forecasting on a per-step ahead basis, has the great advantage of being a gratis and easily applicable alternative to the use of ECMWF-based and NCEP-based ensemble predictions.

3.2. Estimation of the uncertainty for a look-ahead period

In a second stage, the possibility of inform on expected uncertainty for a look-ahead period is considered, by calculating NPRI over a set of successive look-ahead times. The benefits of this temporal integration of uncertainty estimation is assessed by looking at the relation between NPRI_d classes and energy imbalances. Both quantities are calculated over 24 hours (thus for 96 consecutive look-ahead times). While Möhrlen [21] discussed the benefits of considering a larger area when assessing the spread-skill relationship of wind power ensembles, the aim here is to show how skill forecasting can benefit from temporal averaging of spread and skill. Also, adding this temporal component would be relevant for real-world applications, since NPRI_d would be related to levels of energy imbalance over the look-ahead period considered. Note that current methods for uncertainty estimation of wind power predictions only focus on providing pointwise uncertainty estimates (cf. discussion in Pinson *et al* [4]).

The 300 series of ensemble predictions over the 10-month evaluation period are considered. Both NPRI_d values and energy imbalances are calculated for look-ahead times between 0 and 24-hour ahead (day 1), 24 and 48-hour ahead (day 2), etc. NPRI_d values are sorted in 5 equally populated classes. To each of these classes are associated the empirical distributions of related energy imbalances, which contain 60 items each. The same quantities than in the above paragraph are used for summarizing the characteristics of conditional probability distributions (i.e. mean, median, quartiles and 10 and 90% quantiles).

3.2.1. The NPRI_d ability to inform on the expected imbalance level Focus is given first to the same example than that considered in paragraph 3.1, which relates to the use of NPRI_d with

⁴ Here, the spread of the imbalance distributions can be quantified by the inter quartile range, or alternatively by the distance between the quantiles with proportion 0.1 and 0.9. In general, the inter quartile range is preferred, since it consists a more robust measure of the spread of an empirical distribution.

ECMWF-EPW for horizons between 48 and 72-hour ahead. Related conditional probability diagram is depicted in figure 2. As an interpretation of these conditional distributions, one sees for instance that the relative imbalance over day 3 is between 55 and 295% of its climatological level when a NPRI_d value lies in the fifth class. For comparison, this same relative imbalance ranges from 5 to 85% of the climatological value only when NPRI_d belongs to the first class.

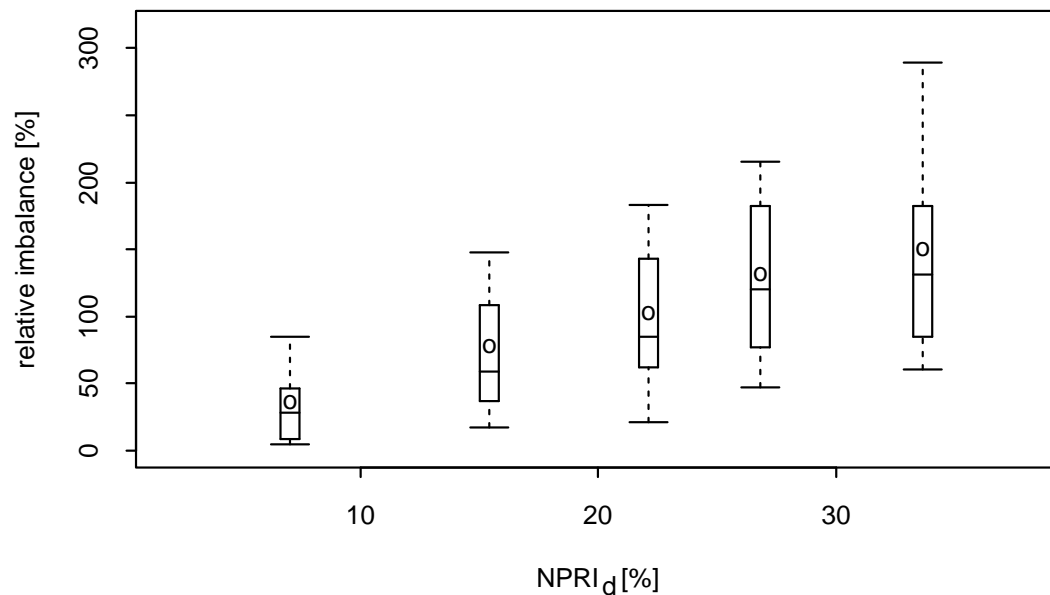


Figure 2. Conditional distributions of energy imbalance given NPRI_d classes. Both imbalances and NPRI values are calculated over day 3 (prediction horizons from 48 to 72-hour ahead). Empirical distributions are made up with 60 elements. Boxplots give the 10% and 90% quantiles (lower and upper tips), the lower and upper quartiles (box bounds), the median (central line) and finally the mean (o).

Similarly to the analysis carried out in paragraph 3.1, the increase of the mean energy imbalance with the NPRI_d class is steady and quasi linear, with mean imbalance levels ranging from 35 to 150% of their climatological value. Therefore, the ability of NPRI_d to be an indicator of expected uncertainty is still valid when considering temporal averaging. In addition, imbalance distributions for every NPRI_d class appear to be sharper than those obtained when using NPRI_h as an uncertainty indicator (cf. figure 1). For instance, the inter quartile range for NPRI class 4 equals 165% in the latter case, while it is only of 105% for the former one. These distributions are sharper first because upper quartiles are at lower level and also because lower quartiles are at a higher level (the same remark is valid for the 10% and 90% quantiles). Temporal averaging of skill smoothes out differences between low and large prediction errors. From a skill forecasting point of view, sharper distributions of expected imbalance are beneficial, since they give more confidence in the estimation of forecast uncertainty.

3.2.2. Comparing the cases for which NPRI_d is calculated with ECMWF-EPW, NCEP-EPW and the lagged average ensembles The comparison between the three types of ensemble predictions is again carried out for day 2. Table 3 gathers some quantiles and the mean of condition

imbalance distributions given the NPRI_d class. If no particular mention, the following remarks are also valid for day 1 and day 3.

Table 3. Characteristics of conditional imbalance distributions given the NPRI_d class. $r^{(\alpha)}$ denotes the quantile with proportion α , while μ relates to the mean. Both NPRI_d and imbalance values are gathered over day 2 (look-ahead times between 24 and 48-hour ahead). Results are for ECMWF-EPW, NCEP-EPW, and the lagged average ensembles respectively. Empirical distributions are made up with 60 elements. Imbalance values correspond to relative imbalances (in % of their climatological level).

(a) - ECMWF-EPW						
NPRI_d class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	3.6	8.6	24.4	48.8	88.3	34.9
2	23.5	41.2	61.4	100.8	171.1	76.9
3	29.7	60.9	94.7	140.3	192.7	112.3
4	67.3	91.1	115.0	164.1	203.0	130.1
5	60.1	102.5	125.6	176.4	222.6	144.6

(b) - NCEP-EPW						
NPRI_d class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	3.5	10.9	24.0	36.9	71.9	35.8
2	23.8	41.0	80.6	110.4	140.7	81.6
3	53.6	70.9	92.4	140.3	189.7	108.9
4	70.5	87.2	118.8	158.0	198.1	129.1
5	61.8	95.6	134.3	181.0	234.5	143.5

(c) - lagged average ensembles						
NPRI_d class	$r^{(0.1)}$	$r^{(0.25)}$	$r^{(0.5)}$	$r^{(0.75)}$	$r^{(0.9)}$	μ
1	7.1	12.9	30.4	49.1	97.9	41.9
2	8.1	44.3	75.8	118.4	170.1	88.9
3	47.4	77.9	96.8	153.2	229.1	116.3
4	29.9	77.4	115.6	154.8	194.9	117.8
5	55.1	88.9	120.9	176.3	195.0	135.5

The ratios between mean imbalances for NPRI_d class 5 and NPRI_d class 1 equal 4.2, 4 and 3.2, for ECMWF-EPW, NCEP-EPW and lagged average ensembles, respectively. The value of the ratio for lagged averaging ensembles is similar to that calculates when focusing on NPRI_h , while it is significantly lower for the two others. This decrease is mainly due to the smoothing of skill, not to a diminution in the ensembles' ability to resolve among situations. In a general manner, ECMWF-EPW and NCEP-EPW are still more skilful for indicating the expected imbalance level. Though, lagged average ensembles gain from the consideration of a temporal component for uncertainty estimation.

Imbalance distributions when considering NPRI_d are much sharper than when considering NPRI_h for the three types of ensembles. The inter quartile range is here between 26 (for NCEP-EPW) and 40% (for ECMWF-EPW) for the first class of NPRI_d values. These values are slightly higher than those in table 3. But for the other NPRI_d classes, it is actually the inverse: the inter quartile range is much lower when imbalance values are sorted depending on NPRI_d values. The reduction of the inter quartile range is up to 50%. Therefore, in terms of skill forecasting, NPRI_d appears to be a better indicator, owing to these sharper distributions.

4. Conclusions

The present investigation on the use of wind power ensemble predictions has revealed their potential for associating skill forecasts to point predictions of wind generation. Focus has been given to the possibility of communicating differently on expected level of prediction uncertainty. Skill forecasts take the form of prediction risk indices. They may be seen as estimates of short-term predictability of wind generation. They thus comprise a comprehensive signal on the confidence forecast users may have in the power predictions provided regardless of the method employed. The prediction risk index NPRI that has been introduced reflects the spread of ensemble members for a single or a set of successive look-ahead times. Various types of wind power ensemble forecasts have been considered: lagged average ensembles (obtained by lagging the ECMWF control forecasts, 5 members), as well as wind power ensembles derived from ECMWF (51 members) and NCEP (11 members) ensemble predictions of meteorological variables. The investigation has been carried out on the case-study of the Tunø Knob wind farm, over a period of 10 months.

The methodology developed has consisted in considering various equally populated classes of NPRI values (more precisely 5 classes), and in establishing their probabilistic relation with energy imbalance levels. It has been shown that for all different types of wind power ensembles considered, NPRI could provide a useful information on expected level of forecast uncertainty. An important point relates to the possibility and interest of defining prediction risk indices for a look-ahead period. They then permit to inform on the level of expected energy imbalance over the period considered. This contrasts with the common providing of pointwise uncertainty estimates. Moreover, an important conclusion is that the gratis alternative of making up wind power ensembles by lagging available point predictions proved to be valuable for estimating the level of expected prediction uncertainty. Considering NCEP-based or ECMWF-based ensembles of wind generation is justified by their better ability of resolving between low and high predictability situations.

Perspectives regarding follow-up studies include: *(i)* a validation of the results on various types of test-cases located in zones with different meteorological characteristics (for which predictability may be more or less easily estimated); *(ii)* further investigation on other possibilities for estimating the disagreement between ensemble members e.g. with categorical measures (mode population and ensemble entropy); *(iii)* a study of other ensemble prediction systems, which may be more appropriate for short-range applications than those considered here; *(iv)* the use of such prediction risk indices in forecast combination or regime-switching methods in order to dampen the risk of large prediction errors.

Finally, a last perspective concerns the real-world utilization of prediction risk indices by end-users of wind power forecasts. As a first step, they can be communicated as a complement to point predictions. This way, forecast users will get used to that information, as a signal on the confidence they may have on the forecasts provided. Then, a second step will be to define how to make alternative decisions (more or less conservative depending on the risk aversion of forecast users) depending on the value of the prediction risk index, and to demonstrate the resulting operational benefits.

Acknowledgments

The results presented have been generated as part of two projects: ‘Wind Power Ensemble Forecasting’ supported by Danish PSO funds (ORDRE-101295/FU2101), and Anemos partly funded by the European Commission (ENK5-CT2002-00665), which are hereby greatly acknowledged. The authors also gratefully acknowledge Elsam Engineering A/S for providing the power data, ECMWF and NCEP for providing the meteorological ensemble data.

References

- [1] Thor S-E and Weis-Taylor P 2002 Long-term research and development needs for wind energy for the time frame 2000-2020 *Wind Energ.* **5** 73-75
- [2] Giebel G, Kariniotakis G and Brownsword R 2003. State of the art on short-term wind power prediction Technical Report Anemos Project deliverable report D1.1 (available online: <http://anemos.cma.fr>)
- [3] Pinson P and Kariniotakis G 2004 On-line assessment of prediction risk for wind power production forecasts *Wind Energ.* **7** 119-132
- [4] Pinson P, Nielsen H Aa, Møller J K, Madsen H and Kariniotakis G 2007 Nonparametric probabilistic forecasts of wind power: required properties and evaluation *Wind Energ.* (to appear)
- [5] Lange M and Focken U 2006 *Physical Approach to Short-term Wind Power Prediction* (Berlin: Springer)
- [6] Palmer T N 2000 Predicting uncertainty in forecasts of weather and climate *Rep. Prog. Phys.* **63** 71-116
- [7] Molteni F, Buizza R, Palmer T N and Petroliagis T 2000 The ECMWF ensemble prediction system: methodology and validation *Quat. J. Royal Met. Soc.* **122** 73-119
- [8] Buizza R, Miller M and Palmer T N 1999 Stochastic representation of model uncertainties in the ECMWF ensemble prediction system *Quat. J. Royal Met. Soc.* **125** 2887-2908
- [9] Toth Z and Kalnay E 1997 Ensemble forecasting at NCEP and the breeding method *Mon. Wea. Rev.* **125** 3297-3319
- [10] Nielsen H Aa and co-authors 2006 From wind ensembles to probabilistic information about future wind power production - Results from an actual application *Proc. IEEE PMAPS Conference*, Stockholm, Sweden
- [11] Pinson P 2006 Estimation of the uncertainty in wind power forecasting PhD Thesis Ecole des Mines de Paris, France
- [12] Toth Z, Zhu Y and Marchock T 2001 The use of ensembles to identify forecasts with small and large uncertainty *Wea. Forecasting* **16** 463-477
- [13] Ziehmann C 2001 Skill prediction of local weather forecasts based on the ECMWF ensemble *Nonlin. Proc. Geophys.* **8** 419-428
- [14] Gritit E P 2004 Redefining the ensemble spread-skill relationship from a probabilistic perspective *NCEP invited presentation*, Camp Spring, Maryland (USA)
- [15] Pinson P, Chevallier C and Kariniotakis G 2007 Trading wind generation with short-term probabilistic forecasts of wind power *IEEE Trans. Power Syst.* (to appear)
- [16] Barker T 1991 The relationship between spread and error in extended range forecasts *J. Climate* **4** 733-742
- [17] Gritit E P and Mass C F 2007 Measuring the ensemble spread-skill relationship from a probabilistic perspective: stochastic ensemble results *Mon. Wea. Rev.* **135** 203-221
- [18] Whitaker J S and Lough A F 1998 The relationship between ensemble spread and ensemble mean skill *Mon. Wea. Rev.* **126** 3292-3302
- [19] Houtekamer P L 1993 Global and local skill forecasts *Mon. Wea. Rev.* **121** 1834-1846
- [20] Moore A and Kleman R 1998 Skill assessment for ENSO using ensemble prediction *Quat. J. Royal Met. Soc.* **124** 557-584
- [21] Möhrle C 2004 Uncertainty in wind energy forecasting PhD thesis University College Cork, Ireland